Preservation metadata initiatives: practicality, sustainability, and interoperability[1]

Michael Day

UKOLN, University of Bath, Bath BA2 7AY, United Kingdom
m.day@ukoln.ac.uk

Abstract. In recent years there have been a range of metadata specifications and frameworks developed to support digital preservation activities. These range from formats that are intended to be specific to certain types of resources (e.g., VERS Encapsulated Objects for electronic records, MPEG-7 for multimedia resources) to generic frameworks based on the information model defined by the Reference Model for an Open Archival Information System (OAIS). Those specifications that exist have been developed from the perspective of a variety of different professional domains and world-views. The paper outlines some of the problems that result from these differing perspectives of preservation metadata initiatives and highlight issues related to their practical implementation and sustainability. A final section considers interoperability issues with reference to the role of metadata registries and Semantic Web technologies.

1. Introduction

In recent years, there has been a growing awareness of the part that metadata can play in supporting the long-term preservation of digital objects. In fact, the key role of metadata in preservation appears now to have been generally accepted, so the focus has now moved on to identifying exactly *what* metadata will be required. For example, the briefing paper prepared for the ERPANET Training Seminar on metadata opened by saying that preserving "the right metadata is key to preserving digital objects" (Duff, Hofman & Troemel, 2003).

This paper will first introduce and attempt to categorise a number of preservation metadata initiatives. These will include the influential Reference Model for an Open Archival Information System (OAIS) and initiatives originating from national and research libraries, the archives and records domain, digitization projects and others. It will then make some comments on the need for practical standards that can be implemented easily, also on the need for sustainability and interoperability. Finally, the paper will propose that the development of metadata registries may help digital

[1] A revised version of a presentation given at the ERPANET Training Seminar on Metadata in Digital Preservation held in Marburg, Germany on the 3-5 September 2003. The proceedings of this event will be published in the series *Veröffentlichungen der Archivschule Marburg*.

repositories manage diverse metadata and could help to facilitate ingest processes and the exchange of metadata and information packages between repositories.

2. Preservation metadata initiatives

Most digital preservation strategies depend to some extent upon the capture, creation and maintenance of appropriate metadata (Day, 2001). Ludäsher, Marciano and Moore (2001) have defined this as all of the various types of data that will allow the re-creation and interpretation of the structure and content of digital data over time. Defined in this way, it is clear that such metadata needs to support a number of distinct, but related, functions. Lynch (1999), for example, has written that within a digital repository, "metadata accompanies and makes reference to each digital object and provides associated descriptive, structural, administrative, rights management, and other kinds of information." The wide range of functions that preservation metadata will be expected to fulfil means that the definition of standards is not a simple task and that most of the currently published schemas are either extremely complex or only attempt to produce a basic framework that can be implemented in different ways. The situation is complicated further by the perception that different kinds of metadata will be required to support different digital preservation strategies or digital information types.

To date, preservation metadata initiatives have tended to originate in one of three distinct contexts, from national and research libraries, the archives and records domain and digitization projects (Day, 2004). The following sections will briefly introduce some of these starting with the highly influential OAIS model.

2.1 The OAIS model

The Reference Model for an Open Archival Information System (OAIS) is an attempt to provide a high-level framework for the development and comparison of digital archives (CCSDS 650.0-B-1, 2002; ISO 14721: 2003). The model aims to provide a common framework that can be used to help understand archival challenges and defines a high-level common language that can facilitate discussion across the many different communities interested in digital preservation. The model defines both a functional model and an information model. The functional model outlines the range of functions that would need to be undertaken by a repository, and defines in more detail those functions described within the OAIS specification as access, administration, archival storage, data management, ingest and preservation planning. The information model defines the broad types of information (or metadata) that would be required in order to preserve and access the information stored in an OAIS-based system.

The OAIS information model defines a number of different Information Objects that cover the various types of information required for long-term preservation. The basic assumption of the model is that all Information Objects are composed of a Data Object -which for digital data would typically be a sequence of bits - and the

Representation Information that would permit the full interpretation of this data into meaningful information. The OAIS model defines four distinct Information Objects:

- *Content Information* - the information that requires preservation
- *Preservation Description Information* (PDI) - any information that will allow the understanding of the Content Information over an indefinite period of time
- *Packaging Information* - the information that binds all components into a specific medium
- *Descriptive Information* - information that helps users to locate and access information of potential interest.

The information model further divides the PDI into four groups, based on the categories of reference, context, provenance and fixity.

The OAIS information model also defines a conceptual structure for Information Packages. These are viewed as containers that logically encapsulate Content Information and its associated PDI within a single Data Object. Information Packages are defined for submission (SIP), archival storage (AIP) and dissemination (DIP). Of these, the Archival Information Package (AIP) can be seen as the most important for digital preservation, as it needs to contain, in principle, "all the qualities needed for permanent, or indefinite, Long Term Preservation of a designated Information Object" (CCSDS 650.0-B-1, 2002, 4-33). Those preservation metadata initiatives that have been informed by the OAIS information model have, therefore, tended to concentrate on the definition of AIPs, and more specifically on the definition of Content Information and PDI.

2.2 National and research libraries

To date, the OAIS model has had most influence on preservation metadata initiatives developed by national and research libraries. Some of these initiatives, although informed by the OAIS model, have essentially been pragmatic responses to the immediate resource management needs of the institution, e.g. the element sets developed by the National Library of Australia (Phillips, *et al*., 1999) and the National Library of New Zealand (2003a; 2003b). Others have been more closely structured on the OAIS model's definition of an AIP, e.g. the specifications developed by the Cedars (Russell, *et al*., 2000) and NEDLIB (Masanès & Lupovici, 2001) projects. These two attempts to define a preservation metadata schema, together with the NLA specification, were taken forward by an international working group convened in 2000 by OCLC Online Computer Library Center and the Research Libraries Group (RLG). The working group produced proposals for Content Information and PDI that were collected together and published as: *A metadata framework to support the preservation of digital objects* (OCLC/RLG Working Group on Preservation Metadata, 2002).

Like the Cedars and NEDLIB specifications, the OCLC/RLG metadata framework uses the OAIS information model as part of its basic structure. Therefore, the recommendation for Content Information includes the Content Data Object (a bit-stream) and as Representation Information, elements that relate to the object itself (e.g., file descriptions, significant properties) or its hardware and software environment (e.g., operating systems). The Provenance Information is organized according to an event-based model, defining generic elements associated with processes that might be carried out on the Content Digital Object, e.g. transformations undertaken at ingest, format migrations, etc. The working group did not envisage that the whole metadata framework would be utilized for each and every object within a preservation system, but that metadata would be implemented at varying levels of specificity. They noted that the elements were not necessarily atomic and that it was "easy to imagine cases where the needs and characteristics of particular digital archiving systems ... [would] require deconstruction of these elements into still more precise components" (OCLC/RLG Working Group on Preservation Metadata, 2002, p. 3).

In 2003, a new group called PREMIS (Preservation Metadata: Implementation Strategies) was convened by the same sponsoring organizations to look at the metadata framework and investigate in more detail the practical aspects of implementing preservation metadata in digital preservation systems, including the identification of 'core' metadata elements.

2.3 Recordkeeping metadata initiatives

In parallel, the archives and records professions have also been investigating what information might be required to support the long-term preservation of digital objects. As might be expected, their primary focus is on records, defined by the ISO (International Organization for Standardization) Records Management standard (ISO 15489:2002) as "information created, received, and maintained as evidence and information by an organization or person, in pursuance of legal obligations or in the transaction of business" (Healy, 2001). Recordkeeping metadata specifications, therefore, tend to have a strong emphasis on the development of systems that ensure the authenticity and integrity of electronic records.

One of the earliest metadata specifications was based on the Business Acceptable Communications (BAC) model developed by the University of Pittsburgh's Functional Requirements for Evidence in Recordkeeping project (known as the Pittsburgh Project). This proposed a metadata structure that would contain a 'handle layer' for basic discovery data while other layers would store information on terms and conditions of use, data structures, provenance, content and the use of the record since its creation (Bearman & Sochats, 1996). Together with other developments, the Pittsburgh Project inspired a series of recordkeeping metadata initiatives, especially in Australia. One of the most interesting of these was the development of a framework known as the Australian Recordkeeping Metadata Schema (RKMS) by a research project led by Monash University. The project, amongst other things, attempted to specify and standardize the whole range of recordkeeping metadata that would be required to manage records in digital environments (McKemmish, *et al*., 1999). The RKMS also was concerned with supporting interoperability with more generic metadata standards like the Dublin Core and relevant resource discovery schemas like the AGLS Metadata Standard. The schema defined a highly structured set of metadata elements conforming to a data model based on that developed for the

Resource Description Framework (RDF) by the World Wide Web Consortium (Manola & Miller, 2004). The schema was designed to be extensible and to be able to inherit metadata elements from other schemas.

There have also been more practical developments. For example, the National Archives of Australia (1999) published a *Recordkeeping Metadata Standard* that defined the metadata that the archives recommended should be captured by the recordkeeping systems used by Australian government agencies. Another significant Australian development was the definition of the Victorian Electronic Records Strategy (VERS) that defined a self-documenting exchange format (the XML-based VERS Encapsulated Object) that permitted the transfer of record content (and metadata) over time (Public Record Office Victoria, 2003). In the UK, the National Archives have developed a metadata standard as part of its definition of functional requirements for electronic records management systems. The standard supports retrieval as well as a range of records management processes, although the 'preservation' section of this standard is still under development (Public Record Office, 2002, p. 28).

2.4 Digitization initiatives

Some of the first projects and initiatives to consider the need for preservation metadata were those involved in the digitization of cultural heritage resources. The early stages of digitisation projects are expensive, and their sustainability depends upon metadata being available to support the long-term management of resources (e.g., Kenney & Rieger, 2000). Much of this metadata is technical in nature, dealing with the attributes of digital images and the production techniques associated with them. To deal with this type of data, in 2002 the US National Information Standards Organization (NISO) issued a data dictionary of "Technical metadata for digital still images" for review as a draft standard for trial use (NISO Z39.87-2002). The data dictionary includes elements that will record detailed information about images themselves (e.g., formats, compression techniques, etc.), the image creation process, quality metrics, and change history (e.g., migrations). No particular encoding of the elements is recommended, although the Network Development and MARC Standards Office of the Library of Congress maintains an XML Schema implementation of it called MIX.

One important recent development has been the development of the Metadata Encoding & Transmission Standard (METS), also maintained by the Library of Congress. This is an attempt to provide an XML Schema for encoding metadata that will aid the management and exchange of digital library objects. A METS document consists of seven sections: the 'METS header,' 'descriptive metadata,' 'administrative metadata,' 'file section,' 'structural map,' 'structural links,' and 'behavior' - some of which are intended to group together all of the files that make up a particular digital object and to link content and metadata to a particular structure. The administrative metadata section is intended to store technical information about the file, as well as information about intellectual property rights held in the resource, the source material, and provenance metadata that records relationships between files and migrations. The modular design of METS means that it can include metadata from 'extension schemas,' e.g. descriptive metadata from the Encoded Archival Description (EAD), the Metadata Object Description Schema (MODS) or Dublin Core, technical metadata from Z39.87, etc. There is also much interest in METS as a potential container for preservation metadata, e.g. as an Archival Information Package or for supporting ingest, dissemination and object exchange functions. For example, Harvard University Library (2001) experimented with METS for defining a Submission Information Package in its Mellon-funded E-Journal Archiving Project.

2.5 Other initiatives

In addition, there are many other metadata standards that contain terms that have relevance to digital preservation. Some of these are format specific or intended for use in particular domains. For example, the MPEG-7 standard (ISO/IEC 15938:2002) is intended to support the management of audio-visual content, and its description schemes can store information about compression methods, data size, access conditions, etc. (Chang, Sikora & Puri, 2001). The IEEE (Institute of Electrical and Electronics Engineers) Learning Object Metadata (LOM) standard (IEEE Std. 1484.12.1-2002) includes elements that describe technical requirements and remarks on installation. It is becoming clear that some of the most important challenges will be making best use of all the relevant metadata that exists in other forms and securing interoperability between the formats used by repositories.

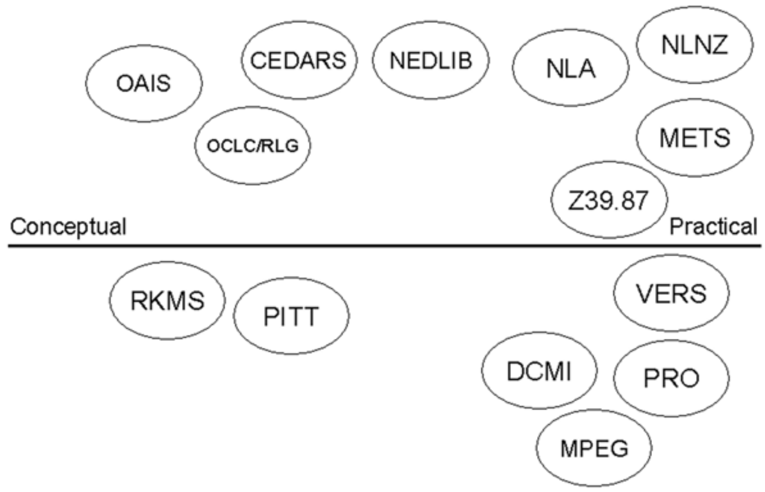
2.6 Characterisation

The plethora of metadata initiatives with relevance to preservation is potentially very confusing. A tentative attempt to categorise initiatives can be found in Figure 1. While there will be many other potential ways of categorisation, the diagram merely places initiatives on a simple continuum from the conceptual to the practical - although this is not intended to be definitive.

On the 'conceptual' side can be found the OAIS information model together with the metadata frameworks closely based on it, also the Australian Recordkeeping Metadata Schema and the Pittsburgh Project's BAC model. On the 'practical' side are the element sets developed by the national libraries of Australia and New Zealand, together with METS and VERS. While not wanting to read too much into this, one may be able to detect a gradual move from the conceptual to the practical. So while the OAIS standard has fulfilled the need for a high-level reference model, the current focus is on developing metadata schemas that can be implemented; a focus also shared by the OCLC/RLG PREMIS working group on implementation strategies.

The need for preservation metadata is now widely acknowledged and various standards have already been developed. There is, however, a need now to acknowledge some more generic issues, including the practicality and sustainability of metadata initiatives and interoperability.

Figure 1: Tentative categorisation of preservation metadata initiatives



3. Practicality

There is an urgent need for metadata schema development to be more securely linked to the practical experiences of preserving digital objects. The tentative categorisation in section 2.6 (above) suggests that things are moving in the right direction, but rather than spend more time developing more frameworks or 'outline specifications,' there is need now to develop schemas that prove the practical value of metadata for supporting long-term preservation. This may mean the development of schemas focused on particular resource types or repository contexts, but the experiences learned from their application can then be filtered back into the development of more generic standards.

Implementation will give us far more *practical* experience of the data that we call preservation metadata. Because of the many roles that it is intended to fulfil - supporting preservation strategies, the integrity of objects, rights management, access control, etc. - preservation metadata schemas tend to be large and complex, but they also risk being based on assumptions that have not been rigorously tested in practice. Solving this problem will be difficult. For example, the developers of the Cedars outline specification (Russell, *et al*., 2000) periodically tested their proposed elements with reference to 'real' objects, but even this process could not help them prove that the schema would enable the successful preservation of these objects.

In short, we have moved beyond the 'proof-of-concept' stage and need far more practical experience of implementing preservation metadata. In turn, these experiences need to feed back into the production and further reiteration of generic standards.

4. Sustainability

It is assumed that the generation and maintenance of preservation metadata will be expensive, although it remains a prerequisite of the successful preservation of digital objects. The difficulty of ensuring digital preservation without metadata may mean that it is ultimately a cheaper and more effective option than the alternative. Chen (2001) has written, "although more semantics in metadata will increase costs, it will minimize human intervention in accessing data; seamless support, transition of stewardship and lifetime maintenance will improve."

Two things, however, may be able to help reduce these costs. The first is that schema developers should be careful about imposing unnecessary costs on the preservation process by ensuring that schemas only define that metadata essential for the long-term preservation of digital objects. As with Duff, Hofman and Troemel (2003), we might characterise this as needing to identify the 'right metadata.'

A second way of reducing metadata costs might be to automatically capture, wherever possible, any metadata that already exist. Hedstrom (2001) has argued that there is a need to identify which aspects of existing metadata standards could be used (or adapted) to support recordkeeping requirements, and the same principle applies to preservation requirements more generally. There is a need for tools that automatically generate some metadata, that can extract it from other schemas on ingest into a repository, and that can capture metadata about preservation processes enacted thereafter (Hedstrom, 2003).

5. Interoperability

The capture and reuse of existing metadata is just one aspect of interoperability that will need to be addressed by digital repositories. Others include managing the growing number of standards currently being developed and implemented and the transfer of metadata or information packages containing metadata to other repositories and services. While the precise way in which future intra-repository co-operation will work remains to be worked out in detail, it seems likely that repositories will need to exchange information packages or metadata with other repositories. One approach to this problem might be to develop standard 'exchange-formats,' possibly based on existing standards like METS. In some domains, e.g. within specific professional communities, it is possible that these formats may emerge as part of the ongoing processes of collaboration and co-operation. In other contexts, it is possible that the exchange of information packages between repositories may become dependent on the sophisticated conversion facilities that could be supported by registries, e.g. of file formats or metadata.

5.1 File format registries

The need for registries of file format information has been recognised for some time. For example, Lawrence *et al*. (2002) argued that there was "a pressing need to establish reliable, sustained repositories of file format specifications, documentation, and related software." Several initiatives are now starting to experiment with these. For example, the Mellon Foundation is funding a research project at University of Pennsylvania on the further development of a Typed Object Model (TOM) that permits the specification of different formats to support their interpretation or conversion. The DSpace repository system includes a 'bitstream format registry,' which is a way of allowing users to precisely identify the format of the resources that they submit to the system (Bass, *et al*., 2002). The UK National Archives has developed PRONOM, an application for managing information about file formats and software applications used to store and render electronic records (Darlington, 2003). Perhaps the most significant of all, is a recent proposal by an ad hoc working group funded by the Digital Library Federation for a global registry of digital formats. Abrams & Seaman (2003) argue that this, if implemented, would "provide an unambiguous and persistent association between an identifier for a format and a set of important syntactic and semantic information about that format, which can be recovered now or in the future in order to facilitate the operation of digital repositories that make use of that format." For example, such a registry would help to identify and validate formats, as well as support OAIS functions like ingest, access and preservation planning.

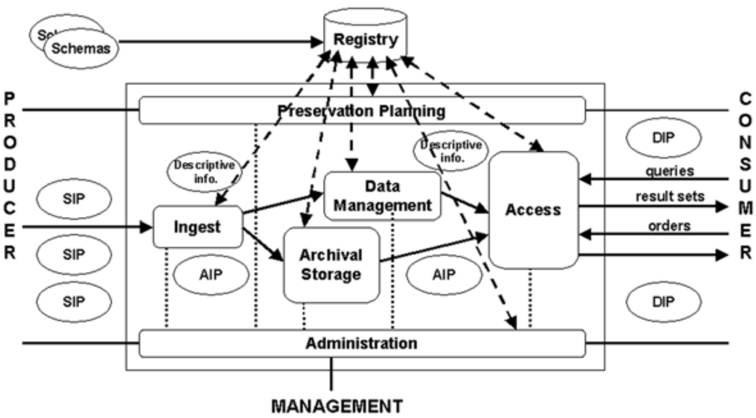
5.2 Metadata registries

Metadata registries have been defined as "formal systems that can disclose authoritative information about the semantics and structure of the data elements that are included within a particular metadata scheme" (Heery, *et al*., 2000). They can include, for example, definitions of terms used, element usage, permitted schemes, and mappings to other standards (Baker, *et al*., 2001). Existing metadata registries take many forms. Some are directories of the data elements used in databases; these are typically based on the ISO/IEC 11179 series of standards. Other types of registry have been designed to support particular types of encoding schemes, formats or subject domains. For example, the XML.org Registry and Repository set up by OASIS (Organization for the Advancement of Structured Information Standards) provides information on XML Schemas and Document Type Definitions (DTDs), with the aim of minimizing their overlap and duplication. The SMPTE (Society of Motion Picture and Television Engineers) Metadata Registry stores authoritative information on data elements (labels) or specifications used for audiovisual content and can also be used for the reconciliation of other metadata schemes (e.g. MPEG-7) within the SMPTE infrastructure (Morgan, 2003). The UK MEG (Metadata for Education Group) Registry provides implementers of educational systems with the means of sharing information about their metadata schemas and supports the re-use of existing schemas (Heery, *et al*., 2002).

In preservation contexts, metadata registries have much in common with file format registries. In the context of a preservation system, registries could provide support for three repository functions (Day, 2003):

- First, like other types of metadata registry, it could support the management function by acting as an authoritative source of information about the metadata terms and vocabularies used within the repository. Wherever possible, metadata would be kept in its original format and the registry would provide information on how it should be interpreted and gives information on its context. The repository could add (or import) information on new metadata schemas when they become available and help manage different versions of the same schema.
- Secondly, once the registry has been populated, it can be used to support the ingest process by providing mappings that could be used to help populate the metadata used by the repository itself. Assuming that the registry maintains mappings from all versions of relevant standards - and that these can be clearly identified - the repository could help automatically populate the metadata that it requires for managing the data and for generating AIPs.
- Thirdly, the mappings maintained within the registry could help support the dissemination or export of metadata or information packages from the repository. Metadata could be collected from the administrative part of the repository and from AIPs, and the registry used to automatically generate selected export formats. While it is highly unlikely that there will ever be a single preservation metadata standard that will be able to be used by all repositories, it may be possible for the different communities to move towards the definition of some kind of standard that might facilitate the exchange of metadata and information packages between repositories.

Figure 2. Registry functions mapped onto the OAIS functional model



Source: adapted from Figure 4-1 in CCSDS 650.0-B-1, 2002

In practice, the registry would also support other functions of the repository, including preservation planning and preservation strategies like migration. Figure 2 is an attempt to map these registry functions on to the OAIS functional model (CCSDS

650.0-B-1, 2002). In the diagram, the registry is modelled as being outside the conceptual Open Archival Information System (e.g. as a service shared between repositories), but it could as easily be seen as part of the system itself, or of its data management function.

Once the functional requirements of metadata registries have been agreed, decisions will then have to be made on how to implement them. For example, a registry could be implemented as a database, in XML, or using Semantic Web vocabularies like RDF.

5.3 Semantic Web technologies

The Semantic Web is a vision of the World Wide Web where the meaning of information can be processed by machines. Berners-Lee and Hendler (2001) stress that the concept of machine-processable documents is not based on artificial intelligence techniques, but "solely on the machine's ability to solve well-defined operations on well-defined data." What this means in practice is that resources are described (or annotated) with semantic markup (metadata) that means that they can be processed by software agents. Semantic Web technologies include the Resource Description Framework (RDF), the RDF Vocabulary Description Language (RDF Schema) (Brickley & Guha, 2004) and ontology vocabularies like the OWL Web Ontology Language (McGuinness & van Harmelen, 2004). RDF provides a simple data model for describing (or representing information about) resources and the relations between them, and an XML-based representation. RDF Schema is a language for describing the properties and classes that are themselves used in descriptions of other resources. Ontology languages like OWL build on RDF and RDF Schema but include additional vocabulary, allowing for the modelling of ontologies, which have been defined as "a set of knowledge terms, including the vocabulary, the semantic interconnections, and some simple rules of inference and logic for some particular topic" (Hendler, 2001).

Semantic Web technologies have many potential applications, e.g. for the integration of data and information or for knowledge management (e.g., Ding, Fensel & Stork, 2003). There has also been much interest in how the Semantic Web might support collaborative and interdisciplinary science or e-science (e.g., De Roure & Hendler, 2004), e.g. for the integration of heterogeneous bioinformatics databases (Staab, 2003; Wroe, *et al*., 2004) or for supporting subject searching in medical databases (Hendler, 2003). In the cultural heritage domains, the Semantic Web is probably of less immediate relevance, partly because institutions currently have so little content that would be usable but also because knowledge representation is, by nature, difficult (Ross, 2003).

Semantic Web languages like RDF have already been used to provide the technical basis of metadata registries. RDF and RDF Schema provide a data model and a vocabulary that can be used for the declaration of metadata schemas. One of the advantages of using RDF is that metadata vocabularies can be managed in a distributed manner. Schema developers can maintain control over vocabularies while registry services can integrate these with other standards, facilitating the reuse of schemas (or parts of them) and respecting diversity between different domains (Heery & Wagner, 2002; Fischer, 2003). This scenario is, however, not problem free. For example, as with the distributed management of ontologies in wider Semantic Web contexts (Maedche, Motik & Stojanovic, 2003), it is not entirely clear how co-dependent metadata schemas would evolve. Current metadata registries based on RDF include the DCMI Registry (Heery & Wagner, 2002) and the MEG Registry, which developed a schema creation tool to create RDF descriptions of metadata vocabularies used to describe learning resources (Heery, *et al*., 2002). The same broad approach was also taken by the EU-funded CORES project, which specifically recognised the need to manage the proliferation of metadata schemas in the digital library and cultural heritage sectors (Heery, *et al*., 2003).

As suggested before, Semantic Web technologies only offer one possible solution to the implementation of metadata registries in preservation contexts. However, Semantic Web technologies do emphasise the importance of accurate modelling and of shared meta-models that can be used across domains. Many current preservation metadata initiatives have derived their underlying model from the OAIS information model. It will be important for future initiatives to investigate how this model might be able to interact with the RDF data model or with other models like the CIDOC Conceptual Reference Model (CRM).

6. Conclusions

This paper has reviewed a range of metadata standards that have been developed to support digital preservation and other related functions. An attempt at characterising these suggests that there is a much greater interest now in metadata solutions focused on practical implementation. Despite this, there remains a need to consider the practicability of those standards being developed and how the production of expensive metadata can be made sustainable, e.g. through the automatic capture of data on ingest and at other stages of the preservation process. Interoperability is another potential problem, and this paper has argued that metadata registries are likely to be a useful way of helping to manage this diverse metadata within a digital preservation system and to facilitate ingest and the exchange of metadata and information packages between repositories and end users. It is possible that Semantic Web technologies like RDF may have a role in supporting these registries.

Acknowledgements. UKOLN is funded by the JISC and the Council for Museums, Libraries and Archives (MLA), as well as by project funding from the JISC, and the European Union. UKOLN also receives support from the University of Bath, where it is based.

References

Abrams, S.L. & Seaman, D. (2003). "Towards a global digital format registry." 69th IFLA General Conference and Council. Retrieved March 3, 2004, from the International Federation of Library Associations and Institutions Web site: http://www.ifla.org/IV/ifla69/papers/128e-Abrams_Seaman.pdf

Baker, T., Dekkers, M., Heery, R., Patel, M., & Salokhe, G. (2001). "What terms does your metadata use? Application profiles as machine-understandable narratives." *Journal of Digital Information*, 2(2), November. Retrieved March 3, 2004, from http://jodi.ecs.soton.ac.uk/Articles/v02/i02/Baker/

Bass, M.J., Stuve, D., Tansley, R., Branschofsky, M., Breton, P., Carmichael, P., Cattey, B., Chudnov, D., & Ng, J. (2002). DSpace -- Internal Reference Specification: functionality. Retrieved March 3, 2004, from the DSpace Federation Web site: http://dspace.org/technology/functionality.pdf

Bearman, D., & Sochats, K. (1996). Metadata requirements for evidence. Retrieved March 3, 2004, from the Archives & Museum Informatics Web site: http://www.archimuse.com/papers/nhprc/BACartic.html

Berners-Lee, T., & Hendler, J. (2001). "Publishing on the Semantic Web." *Nature*, 410, 1023-1024.

Brickley, D., & Guha, R.V., eds. (2004). RDF Vocabulary Description Language 1.0: RDF Schema. W3C Recommendation, 10 February. Retrieved March 3, 2004, from the World Wide Web Consortium Web site: http://www.w3.org/TR/rdf-schema/

CCSDS 650.0-B-1. (2002). Reference model for an Open Archival Information System (OAIS). Retrieved March 3, 2004, from the Consultative Committee for Space Data Systems Web site: http://www.ccsds.org/documents/650x0b1.pdf

Chang, S.-F, Sikora, T., Puri, A. (2001). "Overview of the MPEG-7 standard." *IEEE Transactions on Circuits and Systems for Video Technology*, 11(6), 688-695.

Chen, S.S. (2001). "The paradox of digital preservation." *Computer*, 34(3), 24-28.

Darlington, J. (2003). "PRONOM: a practical online compendium of file formats." *RLG DigiNews*, 7(5), October 15 Retrieved March 3, 2004, from http://www.rlg.org/preserv/diginews/diginews7-5.html#feature2

Day, M. (2001). "Metadata for digital preservation: a review of recent developments." In: *Research and Advanced Technology for Digital Libraries: 5th European Conference, ECDL 2001* (pp. 161-172). Lecture Notes in Computer Science, 2163. Heidelberg: Springer-Verlag.

Day, M. (2003). "Integrating metadata schema registries with digital preservation systems to support interoperability: a proposal." In: *Proceedings of the 2003 Dublin Core Conference* (pp. 1-10). Retrieved March 3, 2004, from the University of Washington Web site: http://purl.oclc.org/dc2003/03day.pdf

Day, M. (2004). "Preservation metadata." In: Gorman, G. E. & Dorner, D. G., eds., *Metadata applications and management* (pp. 253-273). International Yearbook of Library and Information Management, 2003-2004. London: Facet Publishing.

De Roure, D., & Hendler, J. A. (2004). "E-science: the Grid and the Semantic Web." *IEEE Intelligent Systems*, 19(1), 65-71.

Ding, Y., Fensel, D., & Stork, H. -J. (2003). "The Semantic Web: from concept to percept." *ÖGAI Journal*, 22(1). Retrieved March 3, 2004, from the University of Innsbruck Web site: http://homepage.uibk.ac.at/~c703205/download/TheSemanticWeb-OGAI.pdf

Duff, W., Hofman, H., & Troemel, M. (2003). Getting what you want, knowing what you have, and keeping what you need. Retrieved March 3, 2004, from the ERPANET Web site: http://www.erpanet.org/www/products/marburg/marburg.htm

Fischer, T. (2003). "Metadata Registries: Aktueller Stand und zukünftige Entwicklungen." *Zeitschrift für Bibliothekswesen und Bibliographie*, 50(4), 199-204.

Harvard University Library. (2001). Submission Information Package (SIP) specification, v. 1.0 draft. Retrieved March 3, 2004, from the Digital Library Federation Web site: http://www.diglib.org/preserve/harvardsip10.pdf

Healy, S. (2001). "ISO 15489 Records Management - its development and significance." *Records Management Journal*, 11(3), 133-142.

Hedstrom, M. (2001). "Recordkeeping metadata: presenting the results of a working meeting." *Archival Science*, 1(3), 243-251.

Hedstrom, M. (2003). "Research challenges in digital archiving and long-term preservation." NSF Post Digital Library Futures Workshop. Retrieved from the University of Pittsburgh Web site: http://www.sis.pitt.edu/~dlwkshop/paper_hedstrom.html

Heery, R., & Wagner, H. (2002). "A metadata registry for the Semantic Web." *D-Lib Magazine*, 8(5), May. Retrieved March 3, 2004, from http://www.dlib.org/dlib/may02/wagner/05wagner.html

Heery, R., Gardner, T., Day, M., & Patel, M. (2000). *DESIRE metadata registry framework*. Retrieved March 3, 2004, from the DESIRE project Web site: http://www.desire.org/html/research/deliverables/D3.5/

Heery, R., Johnston, P., Beckett, D., & Steer, D. (2002). "The MEG Registry and SCART: complementary tools for creation, discovery and re-use of metadata schemas." In: *Proceedings of the International Conference on Dublin Core and Metadata for e-Communities, 2002* (pp. 125-132) Florence: Firenze University Press. Retrieved March 3, 2004, from http://www.bncf.net/dc2002/program/ft/paper14.pdf

Heery, R., Johnston, P., Fülöp, C., & Micsik, A. (2003). "Metadata schema registries in the partially Semantic Web: the CORES experience." In: *Proceedings of the 2003 Dublin Core Conference* (pp. 11-18). Retrieved March 3, 2004, from the University of Washington Web site: http://purl.oclc.org/dc2003/03heery.pdf

Hendler, J. (2001). "Agents and the Semantic Web." *IEEE Intelligent Systems*, 16(2), 30-37.

Hendler, J. (2003). "Science and the Semantic Web." *Science*, 299, 520-521.

IEEE Std 1484.12.1-2002. IEEE Standard for Learning Object Metadata. New York: Institute of Electrical and Electronics Engineers.

ISO 14721:2003. Space data and information transfer systems -- Open archival information system -- Reference model. Geneva: International Organization for Standardization.

ISO 15489:2002. Information and documentation -- Records management. Geneva: International Organization for Standardization.

ISO/IEC 11179-3:2003. Information technology -- Metadata registries (MDR) -- Part 3: Registry metamodel and basic attributes. Geneva: International Organization for Standardization.

ISO/IEC 15938:2002. Information technology -- Multimedia content description interface. Geneva: International Organization for Standardization.

Kenney, A. R., & Rieger, O. Y., eds. (2000). *Moving theory into practice: digital imaging for libraries and archives*. Mountain View, Calif.: Research Libraries Group.

Lawrence, G.W., Kehoe, W.R., Rieger, O.Y., & Walters, W.H. (2000). *Risk management of digital information: a file format investigation*. Washington, D.C.: Council on Library and Information Resources. Retrieved March 3, 2004, from http://www.clir.org/pubs/abstract/pub93abst.html

Ludäsher, B., Marciano, R., & Moore, R. (2001). "Preservation of digital data with self-validating, self-instantiating knowledge-based archives." *SIGMOD Record*, 30(3): 54-63. Retrieved March 3, 2004, from http://www.acm.org/sigmod/record/issues/0109/SPECIAL/ludaescher8.pdf

Lynch, C. (1999). "Canonicalization: a fundamental tool to facilitate preservation and management of digital information." *D-Lib Magazine*, 5(9), September. Retrieved March 3, 2004, from http://www.dlib.org/dlib/september99/09lynch.html

Maedche, A., Motik, B., & Stojanovic, L. (2003). "Managing multiple and distributed ontologies in the Semantic Web." *VLDB Journal*, 12(4), 286-302.

Masanès, J., & Lupovici, C. (2001). "Preservation metadata: the NEDLIB's proposal." *Zeitschrift für Bibliothekswesen und Bibliographie*, 48(3-4), 194-199.

McGuinness, D. L., & van Harmelen, F., eds. (2004). OWL Web Ontology Language Overview. W3C Recommendation, 10 February. Retrieved March 3, 2004, from the World Wide Web Consortium Web site: http://www.w3.org/TR/owl-features/

McKemmish, S., Acland, G., Ward, N., & Reed, B. (1999). "Describing records in context in the continuum: the Australian Recordkeeping Metadata Schema." *Archivaria*, 48, 3-43. Retrieved March 3, 2004, from the Monash University Web site: http://www.sims.monash.edu.au/ research/rcrg/publications/archiv01.htm

Manola, F., & Miller, E., eds. (2004). RDF Primer. W3C Recommendation, 10 February. Retrieved March 3, 2004, from the World Wide Web Consortium Web site: http://www.w3.org/TR/rdf-primer/

Morgan, O. (2003). "Metadata systems architecture." *SMPTE Motion Imaging Journal*, 112(4), 129-135.

National Archives of Australia. (1999). Recordkeeping metadata standard for commonwealth agencies, v. 1.0. Retrieved March 3, 2004, from http://www.naa.gov.au/recordkeeping/control/rkms/summary.htm

National Library of New Zealand. (2003a). Metadata standards framework -- preservation metadata (revised). Retrieved March 3, 2004, from http://www.natlib.govt.nz/files/4initiatives_metaschema_revised.pdf

National Library of New Zealand. (2003b). Metadata standards framework -- Metadata implementation schema. Retrieved March 3, 2004, from http://www.natlib.govt.nz/files/nlnz_data_model.pdf

NISO Z39.87-2002 AIIM 20-2002. Data dictionary -- Technical metadata for digital still images. Retrieved March 3, 2004, from the National Information Standards Organization Web site: http://www.niso.org/standards/resources/Z39_87_trial_use.pdf

OCLC/RLG Working Group on Preservation Metadata. (2002). *A metadata framework to support the preservation of digital objects*. Retrieved March 3, 2004, from the OCLC Research Web site: http://www.oclc.org/research/projects/pmwg/pm_framework.pdf

Phillips, M., Woodyard, D., Bradley, K., & Webb, C. (1999). *Preservation metadata for digital collections: exposure draft*. Retrieved March 3, 2004, from the National Library of Australia Web site: http://www.nla.gov.au/preserve/pmeta.html

Public Record Office. (2002). Requirements for electronic records management systems: 2, metadata standard. Retrieved March 3, 2004, from http://www.pro.gov.uk/recordsmanagement/erecords/2002reqs/2002metadatafinal.pdf

Public Record Office Victoria. (2003). PROS 99/007 (Version 2) -- Management of electronic records. Retrieved March 3, 2004, from http://www.prov.vic.gov.au/vers/standards/pros9907vers2/default.htm

Ross, S. (2003). "Position paper." In: Geser, G., Kasteren, J. van, Ross, S., & Steemson, M. *Towards a Semantic Web for heritage resources* (pp. 7-11). DigiCULT Thematic Issue 3, May. Retrieved March 3, 2004, from the DigiCULT Web site: http://www.digicult.info/pages/Themiss.php

Russell, K., Sergeant, D., Stone, A., Weinberger, E., & Day, M. (2000). *Metadata for digital preservation: the Cedars project outline specification*. Retrieved March 3, 2004, from the University of Leeds Web site: http://www.leeds.ac.uk/cedars/metadata.html

Staab, S. (2003). "The Semantic Web: new ways to present and integrate information." *Comparative and Functional Genomics*, 4(1), 98-103.

Wroe, C., Goble, C., Greenwood, M., Lord, P., Miles, S., Papay, J., Payne, T., & Moreau, L. (2004). "Automating experiments using semantic data on a bioinformatics Grid." *IEEE Intelligent Systems*, 19(1), 48-55.

Web site references

CIDOC Conceptual Reference Model. Retrieved March 3, 2003, from the Institute of Computer Science of the Foundation for Research and Technology - Hellas (ICS-FORTH) Web site: http://cidoc.ics.forth.gr/

ERPANET (Electronic Resource Preservation and Access Network). Retrieved March 3, 2004, from http://www.erpanet.org/

Global Registry for Digital Format Representation Information. Retrieved March 3, 2004, from the Harvard University Library Web site: http://hul.harvard.edu/formatregistry/

METS - Metadata Encoding & Transmission Standard. Retrieved March 3, 2004, from the Library of Congress Web site: http://www.loc.gov/standards/mets/

MIX - NISO Metadata for Images in XML Schema. Retrieved March 3, 2004, from the Library of Congress Web site: http://www.loc.gov/standards/mix/

PREMIS (PREservation Metadata: Implementation Strategies). Retrieved March 3, 2004, from the OCLC Research Web site: http://www.oclc.org/research/projects/pmwg/

PRONOM. Retrieved March 3, 2004, from the National Archives Web site: http://www.records.pro.gov.uk/pronom/

Typed Object Model (TOM). Retrieved March 3, 2004, from the University of Pennsylvania Web site: http://tom.library.upenn.edu/

XML.org. Retrieved March 3, 2004, from the OASIS Web site: http://www.xml.org/

World Wide Web Consortium (W3C). Retrieved March 3, 2004, from http://www.w3.org/